Explaining and predicting online review helpfulness: The role of content and reviewer-related signals

Michael Siering [a], Jan Muntermann [b,*], Balaji Rajagopalan [c]

[a] *Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany*
[b] *University of Goettingen, Platz der Göttinger Sieben 5, 37073 Göttingen, Germany*
[c] *Northern Illinois University, College of Business, 1425 W. Lincoln Highway, DeKalb, IL 60115, USA*

ABSTRACT

Online reviews provide information about products and services valuable for consumers in the context of purchase decision making. Online reviews also provide additional value to online retailers, as they attract consumers. Therefore, identifying the most-helpful reviews is an important task for online retailers. This research addresses the problem of predicting the helpfulness of online product reviews by developing a comprehensive research model guided by the theoretical foundations of signaling theory. Thereby, our research model posits that the reviewer of a product sends signals to potential buyers. Using a sample of Amazon.com product reviews, we test our model and observe that review content-related signals (i.e., specific review content and writing styles) and reviewer-related signals (i.e., reviewer expertise and non-anonymity) both influence review helpfulness. Furthermore, we find that the signaling environment affects the signal impact and that incentives provided to reviewers influence the signals sent. To demonstrate the practical relevance of our results, we illustrate by means of a problem-specific evaluation scenario that our model provides superior predictions of review helpfulness compared to earlier approaches. Furthermore, we provide evidence that the proposed evaluation scenario provides deeper insights than classical performance metrics. Our findings are highly relevant for online retailers seeking to reduce information overload and consumers' search costs as well as for reviewers contributing online product reviews.

1. Introduction

Online product reviews have become increasingly important in recent years. On the one hand, consumers consider product reviews to obtain information before making their actual purchase decisions [1–3]. On the other hand, online retailers attract consumers by providing a platform that enables customers to exchange their consumption experiences [4].

Given the amount of relevant information provided by online product reviews, a large number of reviews is often beneficial for consumers. However, extensive numbers of product reviews can also create significant information overload for the reader and, hence, high search costs. These costs reduce the use and thereby the value of product reviews [5,6]. To address this problem, online retailers regularly present the most-helpful reviews first. To rank reviews based on helpfulness, several merchants offer their customers the opportunity to vote on whether they perceive a review to be helpful. While the advantage of

* Corresponding author.
E-mail addresses: siering@wiwi.uni-frankfurt.de (M. Siering), muntermann@wiwi.uni-goettingen.de (J. Muntermann), brajagopalan@niu.edu (B. Rajagopalan).

such an approach is that it is based on direct feedback, its limitation is that older reviews (compared to more recent reviews) receive votes over longer periods of time, and more recent reviews have a lower opportunity to gain comparable visibility. Furthermore, it is difficult to rank reviews that have not yet received any votes [5]. Consequently, a priori knowledge about which factors make a review helpful can be the key to highlighting the potential value of reviews. With such knowledge, merchants are able to estimate the helpfulness of online reviews that have not yet been assessed by any consumers. Merchants can then use this estimation to display the most-helpful reviews first.

Recent studies (e.g., [4,5,7]) provide insights into the factors explaining and predicting the helpfulness of online reviews. For example, Mudambi and Schuff [4] show that review depth, review extremity, and product type are factors explaining the helpfulness of online reviews. Other studies have investigated additional factors, such as specific emotions [6,8] or review readability [5,9].

We build upon previous research as well as signaling theory and argue that readers of online product reviews analyze signals related to the review content and signals related to the author of the review. When the reader (i.e., the recipient of the signals) receives the signals, information asymmetry related to the product is reduced, which consequently influences review helpfulness. Thereby, signaling theory

provides a complementary theoretical perspective on review helpfulness, as it allows modeling and exploring the relationship between different signals and review helpfulness. The development of meaningful signal categories within the social commerce context provides a conceptual basis for exploring this relationship [10]. Signaling theory also enables us to extend previous research and to hypothesize on the role of the signaling environment as well as the impact of signaler incentives. We address this research gap as previous studies do not investigate how the presence of other online reviews alters the assessment of review helpfulness and how the provision of reviewer incentives impacts review generation.

When evaluating the predictive power of models on review helpfulness, previous studies make use of classical performance metrics in the form of predictive accuracy [5] and correlations between actual and predicted helpfulness ranking [11]. Nevertheless, neither measure allows for a specific evaluation of how well the most-helpful product reviews can be identified a priori. Our work seeks to build and empirically validate a predictive model for the helpfulness of a product review using a problem-specific evaluation scenario. In this scenario, we assess the predictive performance of the model when online retailers aim at displaying the most-helpful online product reviews first.

In sum, our study builds upon prior work on review helpfulness and provides a comprehensive model to predict the helpfulness of online reviews. Our work contributes to the growing body of knowledge in this domain in the following ways: a) we build upon signaling theory and identify two categories of signals in the context of social commerce (review content-related signals and reviewer-related signals) for studying their relationship with review helpfulness, b) we propose a model to predict review helpfulness and demonstrate the value of these different signals, c) we specifically take into account the role of the signaling environment on signal impact, d) we consider the impact of signaler incentives on the signals sent, and, finally, e) we provide a problem-specific evaluation scenario in order to empirically demonstrate and compare the predictive performance of the proposed model when identifying the most-helpful reviews.

The remainder of this paper is organized as follows. In Section 2, we introduce the theoretical background, providing a basis for our research model, and we derive our research hypotheses and rationale. In Section 3 we present our methodology, including details on dataset generation, textual and statistical analysis, and our empirical evaluation approach. In Section 4, we present our empirical results, apply our novel evaluation scenario, and analyze the predictive performance of the proposed model. Finally, Section 5 concludes the paper and provides a discussion about future research directions.

2. Background and research model

2.1. Review helpfulness and signaling theory

Previous research in the field of review helpfulness builds upon to the economics of information literature and outlines that consumers pursue a purchase decision-making process that aims at reducing uncertainty related to the product. Here, it is argued that review helpfulness is "a measure of perceived value in the decision-making process" and resembles the diagnosticity of the online review related to the reduction of uncertainty [4].

Signaling theory complements this stream of previous research and provides the theoretical foundation of which parties are involved in the field of online reviews, explaining why the different online reviews are contributed and why signals have a differing impact. In this research, we draw upon signaling theory to delineate the relationship between signals conveyed by means of online product reviews and their authors. Furthermore, we build upon signaling theory to explain how these signals are valued by the reader. Extending previous research in the field of review helpfulness, signaling theory also enables us to hypothesize on

the impact of the signaling environment and the impact of signaler incentives.

In short, signaling theory proposes that signals help reduce the information asymmetry between two parties [12]. Here, signals are "in part designed to communicate" and "carry information […] from those with more to those with less information" [12]. The origins of the theory can be found in the labor market [13]. As noted by Spence [14], signals are "activities or attributes of individuals in a market which, by design or accident, alter beliefs of or convey information to, other individuals in the market". Signaling has largely been explored in principal agent situations where one party (agent) possesses more information than the other (principal) [15,16]. Here signals are sent by the agent to reduce the principal's information gap [13,17].

We build upon signaling theory as a theoretical lens in the field of online reviews and identify the key aspects of signaling theory – *signalers, signals, receivers,* and the *signaling environment.* We first briefly outline these key aspects and then relate them to our context of online review helpfulness, where a signaler (reviewer) sends signals to receivers (reader of the review) in a signaling environment (other online reviews) to reduce the information asymmetry related to a product. Consequently, this makes the online review connected to these signals more helpful.

At the core of signaling theory are *signalers* – individuals who generate signals. Typically, these are individuals with insider knowledge about another individual, product, or service. In our context, the insiders or signalers are the product users who have experience with the product and hence have the knowledge that potential users do not yet have access to.

Insiders who have access to private information about a product or a service can choose to divulge this information to the receiver by means of *signals*. Signaling theory distinguishes between two different types of signals, differentiated by a specific level of reliability [10,18]: *assessment signals* and *conventional signals*. Assessment signals require that the signaled quality is possessed and are thus perceived to be reliable. In contrast, *conventional signals* are seen as less reliable. Here, the quality signaled needs not actually be possessed. In the field of information cues displayed in the social commerce context, both assessment signals and conventional signals typically prevail [19].

Generally, a categorization of signals sent in a specific field can be regarded as a research contribution itself [20]. In the context of online reviews, we identify two general categories of signals: *review-related signals* (i.e., signals embedded in the content of a review), and *reviewer-related signals* issued by the social commerce platform. These different signals are clearly observable for users. The costs for sending review-content-related signals are lower, as the signaler can directly include them within the review. In contrast, reviewer-related signals are costlier to obtain, as they require long-term activities or verification by the social commerce platform. Generally, the provision of incentives to the signaler might influence which signals are sent. In the field of online reviews this refers to the provision of free products that have to be reviewed. Nevertheless, the specific influences of signaler incentives have been neglected so far [20].

The *receiver* of the signal is the outsider who is seeking knowledge about the product or service. A key aspect of the receiver is what the individual gains from the signal. In our context, this is straightforward. Receivers are potential users who gain critical knowledge about the product or service they are about to purchase.

Signaling theory also posits that the *environment* might have an influence on the question of how signals are processed, but this specific influence is regarded to be under-researched [20]. In the context of online reviews, further user-generated content in the form of online reviews might be available which might also influence which factors determine review helpfulness. With our study, we close these research gaps.

2.2. Research model

To identify different signal categories, we relied on previous research, which has investigated factors influencing purchase decisions.

We also assume that these factors drive review helpfulness. Here it has been shown that next to the content of a persuasive message, the message source has an influence on the subsequent purchase decision [21–23]. Consequently, within our research model (see Fig. 1), we introduce two categories of signals – *review content-related signals* and *reviewer-related signals*.

With regards to the first category of signals, we recognize that specifically product quality relatedness influences purchase decision making as it reduces the uncertainty towards the product. This can be accompanied by expressed sentiment as well as expressed uncertainty: both variables can emphasize the opinion expressed in the online product review and can therefore influence purchase decision making [24–26] as well as potentially influencing review helpfulness. Regarding reviewer-related signals, we identify user expertise and user non-anonymity as additional factors that may significantly reduce information asymmetry as they are displayed next to the online review and thus may influence the perceived helpfulness.

Furthermore, we account for the actual type of product (i.e., search versus experience good, [27]). *Search goods* comprise a product category about which information can be acquired without difficulty before purchase of the product. In contrast, *experience goods* comprise products that regularly require purchase to enable evaluation of the product and which are more related to image and style than pure functionality. Consequently, the question of how uncertainty related to the product can be reduced by the review's content differs depending on the product type [28], resulting in different moderating effects regarding review-related signals. Finally, we also account for different control variables, which have already been shown in previous research to influence review helpfulness.

In the following sections, we elaborate on how signals embedded in the review content and signals related to the characteristics of the review author can explain review helpfulness for both experience and search goods. Furthermore, we hypothesize on the influence of the information environment as well as signaler incentives. The developed model can thereafter be evaluated and applied in order to predict review helpfulness.

2.2.1. Review content-related signals

As review content-related signals, we assess statements related to *product quality*, as they reduce the uncertainty towards the product. Furthermore, product quality relatedness might be accompanied by *strength of review sentiment* and *review uncertainty*, which both emphasize the product evaluation expressed in the review. These signals represent conventional signals, as the signaler can directly influence them.
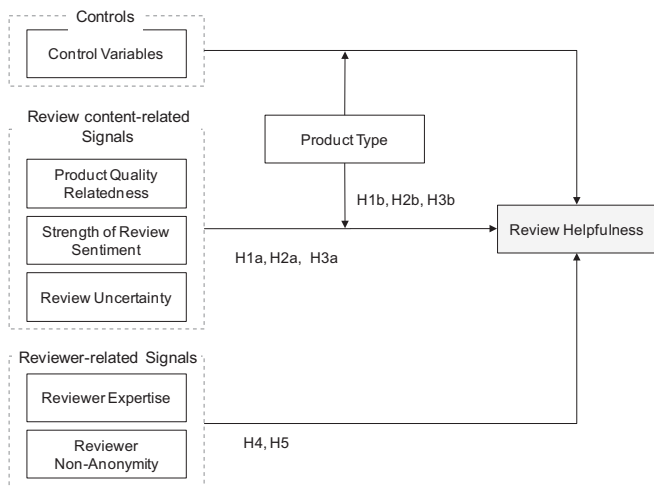

Fig. 1. Research model of predictors of review helpfulness.

2.2.1.1. Product quality relatedness. Based on product characteristics, (missing) functionalities and features, consumers make assessments about actual product quality. Zeithaml [24] defines product quality as a "consumer's judgment about a product's overall excellence or superiority". From a consumer perspective, information about product quality represents a valuable kind of information and can form consumer attitudes and shape buying intentions [29,30]. However, information about product quality, compared for example to product prices, is not easily obtained by customers [27].

A review providing such information helps to reduce the information asymmetry between the actual user of a product (i.e., the reviewer or signaler) and the reader of the review (i.e., the potential buyer or receiver of the signal). We therefore hypothesize on a positive association between the volume of statements relating to product quality and review helpfulness:

H1a. A higher volume of signals related to product quality is associated with a higher review helpfulness rating.

We also consider the product type as a factor moderating this relationship. Since the individual characteristics and features of search goods can be evaluated more easily compared to experience goods [27,31], we expect that information about product quality provided in a product review is more helpful for customers when they assess experience goods. Against this background, we posit the following hypothesis:

H1b. The impact of the volume of signals related to product quality on review helpfulness is moderated by product type. This relationship is stronger for experience goods relative to search goods.

2.2.1.2. Strength of review sentiment. Emotions have major effects on decision-making. The sentiment expressed in online reviews represents relevant information, since it affects consumers' purchase decisions [32]. In general, sentiment has been defined as "attitude, thought, or judgment prompted by feeling" [33].

If the sentiment expressed in the review underlines the reviewer's product evaluation (i.e., a positive evaluation is underlined by positive sentiment, and vice versa), the evaluation of the product becomes clearer. Thus, the review should be perceived as more helpful by consumers because it signals more clearly whether the product should be considered or not [34]. Consequently, the strength of the signaled sentiment is one key for reducing the information asymmetry between the sender and the receiver. Against this background, we posit:

H2a. A higher signaled strength of sentiment (positive or negative) is associated with higher review helpfulness ratings.

We also consider the product type as a factor moderating the relationship between strength of sentiment and review helpfulness, whereas less-provoking opinions with a lower strength of sentiment can be assumed to be more helpful in the case of experience goods, as they are perceived as less offending by readers who disagree with the expressed opinion. As follows, we posit that the strength of sentiment has a less positive impact on review helpfulness in the case of experience goods:

H2b. The relationship between a higher signaled strength of sentiment (positive or negative) and review helpfulness is moderated by product type. Experience goods exhibit a weaker link between the strength of sentiment and review helpfulness compared to search goods.

2.2.1.3. Review uncertainty. The positive effect of expressed certainty on individual perceptions is well documented in the literature. In this context, (un)certainty is defined as "the degree to which an individual is [not] confident that his or her attitude toward an object is

correct" [35]. While certainty that an expressed opinion is correct is important regarding its persuasive power [36], such confidence also has a positive effect in purchase decisions [37]. In the context of product reviews, expressed certainty becomes a relevant issue, because readers expect definite assessments of relevant product features [26]. The opposite can be expected if uncertainty is expressed: When a product review is signaling high uncertainty, the information provided within the review is perceived as less helpful for reducing the information asymmetry. Against this background, we propose:

H3a. A higher level of signaled uncertainty is associated with a lower rating of review helpfulness.

Subjectivity plays a major role when evaluating experience goods, since individual perceptions become relevant [27]. Thus, higher levels of uncertainty in case of experience goods imply a reduced confidence in the review and thus reduce the lowering of information asymmetry. Nevertheless, the review is still valuable, given the little information available about the product otherwise. In the case of experience goods, and when there is a disagreement between the author and the reader about the product, higher uncertainty expressed in the product review may be recognized as more useful. For experience goods, we expect a smaller negative effect of uncertainty expressed in the product review. We therefore hypothesize the following on this moderating effect:

H3b. The relationship between level of signaled uncertainty and review helpfulness is moderated by product type. The relationship is weaker for experience goods than for search goods.

2.2.2. Reviewer-related signals

Regarding reviewer-related signals, we identify reviewer expertise and reviewer non-anonymity as additional signals that may significantly affect the perceived helpfulness of a review, as they are displayed next to the online review and indicate credibility and validity of a reviewer's qualities [38]. Accordingly, these factors may also affect the helpfulness of product reviews.

2.2.2.1. Reviewer expertise. Previous research in the field of advice acceptance has shown that the expertise of the recommending users is important when purchase decisions are made [22]. In the context of online product reviews, the expertise of a specific reviewer is displayed in the form of the rank that the corresponding reviewer has achieved [39]. The rank often mainly incorporates the number of online product reviews contributed. Thereby, with an increased number of contributions, a reviewer's experience writing helpful online product reviews can be expected to increase. Additionally, reviewers receive feedback from other consumers, which also leads to increased expertise [40]. Finally, if a social commerce website offers customers the opportunity to evaluate reviewers and then incorporates this information in the user rank, this also contributes to a reviewer being perceived as an expert [41].

Therefore, a high user rank signals to other users that a reviewer has gained expertise during his activities on the social commerce platform. In the context of signaling theory, the display of the previous reviewer experience in the form of a user rank resembles an assessment signal (i.e., the corresponding reviewer has contributed a certain quantity of helpful online product reviews), which has improved the corresponding rank. The rank assigned by the platform can also be interpreted as a seal [42], which reduces the uncertainty perceived by the consumer. As such, this signal can also be assumed to influence review helpfulness. Consequently, we hypothesize the following:

H4. The source of the review is associated with review helpfulness. A higher level of signaled expertise is associated with higher ratings of review helpfulness.

2.2.2.2. Reviewer non-anonymity. In the social commerce context, online product reviews can be posted anonymously, which might reduce their credibility because visitors to social commerce websites must assess the review's credibility from other cues (i.e., the review's content or other reviewer-related aspects) [43]. In addition, due to reputation concerns, fake reviews are typically written by reviewers not disclosing their real name [44].

Reviewers posting their reviews have the flexibility to decide under which user name these online product reviews are published. A real name can be seen as an assessment signal because in the case of online retailers, such as Amazon, the real name is verified by the retailer and is generally perceived as reliable. Because online retailers most often need a customer's real name to be able to ship the ordered products, verification of the displayed user name provides a reliable assessment signal if the online retailer displays a badge next to the user name. In this context, Forman et al. [2] provide first evidence of the relationship between information disclosure and review helpfulness, but they do not directly focus on user non-anonymity.

Posting unreliable online product review content connected with a reviewer's real name could potentially damage an individual's online reputation. Reviewers active under their real names can be assumed to place more emphasis on writing reliable online product reviews, which can be expected to be more helpful. Consequently, signaling the real name of the reviewer is likely to have a positive influence on the review's reliability, its impact on information asymmetry, and review helpfulness. Against this background, we hypothesize the following:

H5. The source of the review is associated with review helpfulness. If reviewers signal their identity, their reviews are associated with higher ratings of review helpfulness.

2.2.3. Control variables

To demonstrate the robustness of our observations and to compare the influence of review content- and reviewer-related signals with existing research, our proposed research model also takes into account several control variables.

Review depth, review readability, and review extremity are included as relevant factors influencing review helpfulness, as they represent the way the content-related signals are transmitted. Furthermore, we include the moderating effect of the product type [4]. As additional control variables, we include the age of the online product review, the squared review depth to represent possible information overload [45], control variables for the different Amazon product categories, monthly dummy variables to control for seasonality effects and the number of helpfulness votes received. We do not add a control variable for a potential price effect, as the price range within a product category is quite similar but there is price variance across product categories. Consequently, adding a specific variable accounting for high and low product prices would result in high similarity with the product category controls, which would thus lead to multicollinearity.

2.3. Influence of the signaling environment

On social commerce platforms, a varying amount of online reviews can be available related to a specific product. Consequently, consumers have the opportunity to consider different reviews in order to grasp the consumer perception of the product and thus to reduce product-related uncertainty. Related to the signals sent by means of online reviews, it is apparent that different online reviews may compete with each other for the consumer's attention, which might also influence whether specific signals embedded in the online review are perceived to be helpful or not [20,46,47]. In an environment consisting of various online reviews, it can be assumed that specific signals are necessary to make an online review helpful [48], whereas in an environment consisting of few online

reviews, it can be assumed that already the existence of the online review itself is helpful to consumers. Consequently, we hypothesize:

H6. The influence of signals sent on review helpfulness depends on the signaling environment.

2.4. Influence of signaler incentives

In some cases, reviewers are incentivized to publish online reviews on social commerce platforms. Online retailers such as amazon.com offer selected reviewers the opportunity to receive products at no charge on condition that a review about the product received is provided. If a reviewer receives an incentive, he provides more comprehensive online reviews [49] as the effort spent for contributing the online review increases [50]. In this context, it can also be assumed that the reviewer will specifically take care of sending signals which make the review helpful to consumers, so that the reviewer will also receive additional free products in the future. It can thus be assumed that these incentives have an influence on the signals sent by means of the online review. Hence, we posit:

H7. Reviewers who receive incentives for providing online reviews send more signals than reviewers who do not receive incentives.

3. Research methodology

3.1. Research process

To predict the helpfulness of online product reviews, we follow a structured knowledge discovery process [51]. The research process applied is outlined in Fig. 2 and consists of five steps.

We first acquired online product reviews from Amazon's website and cleaned the data by removing duplicate entries. Thereafter, we pre-processed the data by performing automated content analysis and transformed the data by operationalizing the variables used in our study. Finally, we evaluated our research model by means of statistical analysis and then predicted review helpfulness by means of our predictive evaluation scenario and assessed the predictive performance of our model.

3.2. Data acquisition, cleaning and selection

We collected our product review data from Amazon's website. For the two product types, we collected data on different product categories and selected the 100 best-selling products for each category. Our data is comprised of many reviews related to search goods (Camera & Photo, Computer Printers and Cordless Telephones) and experience goods (Music and PC-compatible Games). The product categories were selected following the guidance of Mudambi and Schuff [4]. However, we did not include MP3 players as experience goods. Next to products with prevailing experience good characteristics (such as the iPod), this category also contain products with prevailing search good characteristics (such as no-name MP3 players). This would therefore make it difficult to properly assign the product category to a single product type.

For each single product, we collected the related reviews and the reviews' metadata. Specifically, we downloaded the review text, star rating, the number of helpful votes it received from its readers and the total number of readers who provided a rating related to the review. Furthermore, we acquired reviewer-related data (i.e., the rank and whether a reviewer reveals his/her real name). Overall, we focused on product reviews that received at least ten votes in total to ensure the reliability of the results. Focusing on at least ten votes ensures that the corresponding helpfulness measure is fine-grained enough to facilitate a proper analysis, as the maximum step size of the helpfulness measure is 0.1. Nevertheless, to control for the robustness of our empirical results, we also performed our analysis with a sample of reviews having received at least five (maximum step size of 0.2) or at least fifteen votes (maximum step size of 0.067). The results remained robust in these cases. Because Amazon sometimes lists the same product multiple times on the best-sellers page (e.g., different product configurations such as colors) but displays identical product reviews for the different product configurations, we only collected each product review once to avoid redundancies within our dataset. To avoid overfitting during the practical evaluation of the model, we split the sample into two random subsamples, with 2/3 of the data used as a training sample and 1/3 of the data used as a holdout sample [52]. For each review in the holdout sample, we then predicted the helpfulness by means of our proposed model built upon the training sample.

3.3. Pre-processing of online reviews

To be able to analyze the impact of review content on review helpfulness, we performed a content analysis. This data analysis procedure extracts information from texts by reducing the whole amount of content to "manageable bits of data" [53]. Having been applied in different fields, most importantly in the field of psychology to draw conclusions on message authors [54] as well as individuals who communicate, content analysis generally consists of different methodological procedures that ensure an objective processing of texts [55].

Content analysis can encompass manual coding and different automated analysis procedures. In automated content analysis, two broad strategies can be distinguished [54]: approaches based on machine learning and dictionary-based approaches. Machine learning-based approaches require appropriate training data to train classifiers that are utilized for further document classification [56]. During the generation of such training data, issues related to inter-coder reliability can emerge. Furthermore, the manual labeling of documents is more time-consuming and error prone compared to the application of standardized and extensively validated dictionaries.

In contrast, dictionary-based content analysis is free of problems with inter-coder reliability, as standardized dictionaries are used [53]. Additionally, dictionary-based content analysis has been shown to be robust [54,55] and does not consume the same high amount of time as manual coding [54]. Finally, the outcome of automated content analysis can be replicated without much effort, and the dictionaries applied are often accessible to the public, which makes processing of texts understandable [57].

During dictionary-based content analysis, text documents are mapped to specific categories (e.g., 'positive' or 'negative') contained in the dictionary [53]. In the following, we apply the dictionary of the General Inquirer (GI) [58,59]. The GI is a well-established framework for content analysis (see [58]). Applying the GI has several advantages, such as the validation of the dictionary as well as the resulting standardized classifications [60]. Kelly and Stone [61] found that >90% of the classifications of such a dictionary-based approach were made correctly.
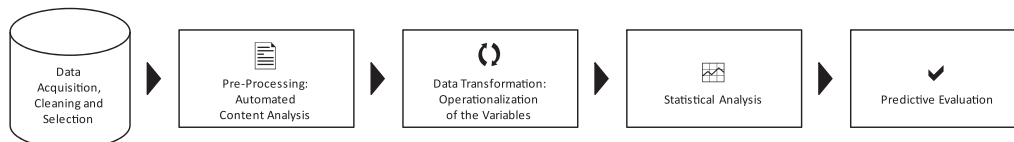
Data Acquisition, Cleaning and Selection ▶ Pre-Processing: Automated Content Analysis ▶ Data Transformation: Operationalization of the Variables ▶ Statistical Analysis ▶ Predictive Evaluation

Fig. 2. Research process followed.

Furthermore, the comprehensiveness with regards to categories covered by the GI allows us to assess three different categories (related to quality, sentiment, and uncertainty) using only a single dictionary.

We automatically analyzed the online review content of the different items in our data set. First, we extracted product quality by relying on the GI, whereas we focused on the amount of terms in the category of "Quality" (exemplar words are "quality", "secure", "stable" or "weak"). The category contains "words indicating qualities or degrees of qualities which can be detected or measured by the human senses".[1] We then related this amount of terms to the overall number of words within the review. To measure the sentiment of the review, we focus on the General Inquirer categories "Positiv" and "Negativ" (exemplar terms are "excellent" and "great" or "awful" and "unhappy") and account for negations. To measure whether the sentiment expressed in the review underlines the review's overall rating, we calculate the variable DirectionalSentiment. Thereby, we compare a review's star rating with the average star rating regarding a specific product. If the star rating is greater or equal to (smaller than) the average star rating, the review is more positive (negative) than the average review. Thus, directional sentiment is defined as the ratio of positive (negative) terms related to the total number of terms within the review and thus measures how strong the sentiment expressed supports the product evaluation. Finally, we determine the review uncertainty score by considering the "If" category, which denotes "feelings of uncertainty, doubt and vagueness"[2] (exemplar terms are "almost", "may" or "seem"). Table 1 provides examples for different statements contained in online product reviews that are representative of the different content categories.

3.4. Variable operationalization

For measuring the variables of interest, we apply content analysis to extract our measures from the product review contents. Table 2 summarizes the different variables used in our further analyses.

Concerning the review content, we measure the different variables as already outlined within the previous section. Regarding reviewer-related signals, we measure expertise by whether a reviewer has achieved a certain rank. In this context, the rank considers the quantity and quality of the previous reviews posted, also weighted by the time that has passed since the reviews have been posted [62]. Furthermore, we take into account a badge displaying whether a reviewer uses his real name as a user name.

Additionally, we determine different control variables, following Mudambi and Schuff [4]. We control for product type (search versus experience good), review depth, reflecting the amount of reasoning in the review (amount of words), and the number of people who voted on review helpfulness. According to [35], review extremity is measured by measuring the absolute difference of the review's star rating and the average product's star rating. This also accounts for possible rating inflations [63]. Consequently, we measure how a specific review deviates from the typical review. Furthermore, we included dummy variables for the different product categories, such as Computer Printers or Cordless Telephones. To control for seasonality effects, we also added monthly dummy variables. Finally, we determine the helpfulness of a review by calculating the ratio of the number of helpful votes to the total number of votes. Thereby, a review is considered to be helpful if it supports the purchase decision making process [4].

3.5. Tobit model

We apply a Tobit regression analysis to examine how the different variables influence review helpfulness [4,6,9]. Compared with Ordinary Least Squares regression, Tobit regression can be regarded as

---

[1] Source: www.wjh.harvard.edu/~inquirer/homecat.htm
[2] Source: www.wjh.harvard.edu/~inquirer/homecat.htm

Table 1
Example statements on different content categories.

Category	Example
Quality	"Works well, very clear, crisp image quality on the screen."
	"This is a really nice product, light and easy to carry."
Positive sentiment	"I love this album! One Direction rocks! … Best. Band. EVER!"
	"Excellent sound and quality… Good deal."
Negative sentiment	"Awful. Just plain old awful. Shell out a bit more for a better model, I beg you. Save yourself the grief."
	"Bad release, bad "fixing" patches, bad forum control, bad items, just a general bad waste of time."
Uncertainty	"Purchased two of these for gifts. They both seem to like them."
	"The product I bought from Amazon as a Warehouse Deal seems to work perfectly."

advantageous for several reasons: within this study, we have a censored dependent variable [4]. Here, review helpfulness has a lower and an upper limit: review helpfulness is defined as zero if no reader votes that the review is helpful. If all readers perceive the review to be helpful, the review helpfulness would have a value of 100%. Values below zero and above 100% are not possible. Second, not every reader votes on helpfulness, so a selection problem exists [4].

The resulting regression equation is shown in Eq. (1). Therein, we consider the different independent variables, the hypothesized moderating effects, the different control variables (ProductType, Extremity, Depth, Depth^2, Age, TotalVotes, Readability, Product Controls, Monthly Controls) and, following Mudambi and Schuff [4], the subsequent interactions (Extremity × ProductType, Depth × ProductType, and, due to consistency, Depth^2 × ProductType).

$$
\begin{aligned}
Helpfulness = {} & Constant + \beta_1 Quality + \beta_2 DirectionalSentiment \\
& + \beta_3 Uncertainty + \beta_4 (Quality \times ProductType) \\
& + \beta_5 (DirectionalSentiment \times ProductType) \\
& + \beta_6 (Uncertainty \times ProductType) \\
& + \beta_7 UserRank + \beta_8 RealName + \beta_9 Controls + \varepsilon
\end{aligned}
\tag{1}
$$

3.6. Predictive evaluation

To predict review helpfulness and to evaluate the practical relevance of our model, we assess whether the model achieves the main goal: identifying the most-helpful online product reviews to display them first. This should correspond to the main goal of online retailers. Consequently, we are interested in the accuracy of the proposed model, taking into account a special requirement: the reviews displayed as most-helpful must be assessed properly, whereas the correct assessment of review helpfulness for less-helpful ranks is not that important.

Therefore, we follow a multi-step evaluation procedure. First, for each online product review in the holdout sample, we estimate the review helpfulness with our proposed model. Second, we rank the different reviews according to their estimated helpfulness to obtain the most-helpful product reviews. Third, we determine whether and how severely the estimated rank deviates from the actual rank based on the users' helpfulness votes. This is of special importance, as online retailers aim at displaying a certain number of most-helpful reviews first. Fourth, when comparing the estimated rank of each review with the actual rank of the review based on the readers' helpfulness assessment, we calculated the mean absolute error MAE [64] (i.e., the mean difference between the estimated and the actual rank), to evaluate the proposed model.

This evaluation methodology is of special practical relevance and extends previous research, which has mainly considered binary helpful versus not helpful classifications [5] or simply the ranking of online

Table 2
Operationalization of Independent (IV) and Dependent Variables (DV).

Variable type	Research hypothesis		Variable	Operationalization
IV	Review content-related signals	H1: Product quality	Quality	Ratio of 'Quality'-terms according to the GI related to the entire number of words.
		H2: Review sentiment	DirectionalSentiment	Defined depending on a review's star rating related to the average star rating of the product: If a review's star rating is greater than or equal to the average star rating of the product: Ratio based on GI-Category 'Positiv' related to the total number of words. If a review's star rating is smaller than the average star rating of the product: Ratio based on GI-Category 'Negativ' related to the total number of words.
		H3: Review uncertainty	Uncertainty	Ratio of 'If'-terms according to the GI related to the entire number of words.
	Reviewer-related signals	H4: Reviewer experience	Rank	Rank of the reviewer, measured by the logarithm of the Amazon Rank.
		H5: Reviewer non-anonymity	RealName	Dummy variable, 1 if real name badge is displayed, 0 otherwise.
	Controls		ProductType	Dummy variable, 1 for experience goods, 0 for search goods.
			Extremity	Absolute value of Star Rating minus Mean Rating.
			Depth	Number of Words.
			Age	Age of the Product Review (Days since January 1, 1960; converted by Stata).
			TotalVotes	Number of Votes.
			Readability	Automated Readability Index.
			Product Controls	Dummy variables for the different Amazon product categories.
			Monthly Controls	Dummy variables for the month the review is written.
DV			Helpfulness	Ratio of Helpful Votes to TotalVotes.

product reviews [11]. Instead, our evaluation methodology addresses the identification of the most-helpful reviews from a holistic perspective. To be able to compare our practical evaluation scenario related to the predictive performance of the proposed model with previous evaluation approaches, we also calculate the classification accuracy for a binary helpfulness prediction [5] and the correlation of the predicted ranks [11].

4. Empirical analyses

4.1. Descriptive statistics

The dataset analyzed within this study encompasses 12,330 online reviews. In Table 3, we show that 4067 reviews address search goods and that 8263 reviews address experience goods.

First, we investigate the difference between online product reviews related to search and experience goods. Therefore, we apply the Wilcoxon-signed-rank test to test whether the medians of the variables are equal. Related to statements about product quality, Table 3 provides the insight that reviews about search goods consist of a significantly larger quantity of signals related to product quality. Furthermore,

reviews about experience goods express higher uncertainty and contain a more positive directional sentiment compared to search goods. One explanation for this observation is that the reviewer's evaluation of experience goods is regularly more subjective, since authors are less certain.

Concerning the reviewer-related signals, we observe that more experienced reviewers focus on search goods. In contrast, there are more reviews written by non-anonymous reviewers about experience goods than about search goods. Considering the control variables, reviews on experience goods are more extreme than reviews on search goods. The average star rating of search goods is 3.3541 and the average star rating on experience goods is 2.4786. Finally, we the data suggest that online reviews about search goods show a higher helpfulness rating than reviews about experience goods [4].

Furthermore, we analyze the variable correlations to ensure that our results are not driven by multi-collinearity. As shown in Table 4, there are only very low correlations between the independent variables. Interestingly, we observe a negative correlation when considering review helpfulness and review extremity: when the star rating of the review deviates from the average star rating, this causes a decrease in review

Table 3
Descriptive statistics.

Variable	Full Sample				Reviews on Search Goods				Reviews on Experience Goods				p-value
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	
Quality	0.0183	0.0181	0.0000	0.5000	0.0250	0.0217	0.0000	0.5000	0.0149	0.0150	0.0000	0.3333	0.000
DirectionalSentiment	0.0625	0.0467	0.0000	1.0000	0.0553	0.0408	0.0000	0.5000	0.0660	0.0489	0.0000	1.0000	0.000
Uncertainty	0.0175	0.0151	0.0000	0.5000	0.0163	0.0149	0.0000	0.5000	0.0181	0.0152	0.0000	0.2000	0.000
Rank	13.6203	2.3872	1.3863	16.2663	12.6954	2.5553	2.0794	16.2648	14.0756	2.1592	1.3863	16.2663	0.000
RealName	0.3635	0.4810	0.0000	1.0000	0.3474	0.4762	0.0000	1.0000	0.3714	0.4832	0.0000	1.0000	0.009
Extremity	1.4817	0.9597	0.0000	3.8077	1.4198	1.0344	0.0000	3.7527	1.5121	0.9192	0.0000	3.8077	0.000
Depth	249.453	324.902	1	6198	287.083	327.374	2	3928	230.932	322.088	1	6198	0.000
Age	18,610.0	662.8	14,497.0	19,296.0	18,789.6	433.40	14,941.0	19,292.0	18,521.7	734.4	14,497.0	19,296.0	0.000
TotalVotes	38.041	131.671	10	9297	46.985	120.125	10	3020	33.638	136.790	10	9297	0.000
Readability	10.746	8.927	−6.300	196.770	11.066	9.636	−4.230	175.714	10.589	8.553	−6.300	196.770	0.000
Helpfulness	0.6229	0.2976	0.0000	1.0000	0.7295	0.2959	0.0000	1.0000	0.5704	0.2842	0.0000	1.0000	0.000
n	12,330				4067				8263				

Table 4
Variable correlations.

	Variable	1	2	3	4	5	6	7	8	9	10	11
1	Quality	1.00										
2	DirectionalSentiment	0.11	1.00									
3	Uncertainty	−0.02	−0.02	1.00								
4	Rank	−0.11	0.08	0.00	1.00							
5	RealName	−0.01	0.00	0.01	−0.05	1.00						
6	Extremity	−0.03	0.03	0.01	0.31	−0.05	1.00					
7	Depth	0.03	−0.12	0.05	−0.33	0.02	−0.14	1.00				
8	Age	0.05	−0.01	0.00	−0.15	−0.16	0.05	0.10	1.00			
9	TotalVotes	0.01	−0.03	0.01	−0.15	0.02	−0.02	0.16	−0.04	1.00		
10	Readability	−0.01	−0.10	0.01	−0.07	−0.01	−0.04	0.13	0.03	0.03	1.00	
11	Helpfulness	0.09	−0.17	−0.04	−0.52	0.03	−0.54	0.21	−0.04	0.07	0.04	1.00

helpfulness. Furthermore, our results show that the GI categories proposed in our study are nearly independent (i.e., terms included in a specific word list are not often contained in other word lists). Furthermore, there is a very low correlation of the rank and the variables representing the content-related signals, which shows that there is no spillover effect of the rank on the review content-related variables.

4.2. Empirical results

4.2.1. The drivers of review helpfulness

The results of our Tobit regression analysis are shown in Table 5, which presents the results of the base model considering the control variables, a second model adding the review content-related signals and a third model adding the reviewer-related signals.

Focusing on review content-related signals, we find support for H1a, which focuses on the effect of product quality-related statements on the helpfulness of online reviews at a 0.1% level of significance. Consequently, if an online review focuses on product quality, then its helpfulness is increased. However, H1b is not supported. There is no significant moderating effect of Quality and ProductType.

We also find support for H2a, as a high strength of directional review sentiment leads to an increase of review helpfulness perceptions. This effect is significant at a 0.1% level. Related to H2b, we observe that in the case of experience goods, the impact of a high strength of directional review sentiment is lower. We thus corroborate the moderating effect of the product type (significant at a 0.1% level). For experience goods, high sentiment strength even has a negative overall influence on review helpfulness.

Focusing on H3a, we test the negative impact of uncertainty-related statements on the helpfulness of online reviews. H3a is supported, as shown by the negative coefficient, which is significant in any case at a 5% level. If an online review contains terms expressing uncertainty because its author is not convinced of the review, the review receives a lower helpfulness assessment. Related to H3b, we do not observe a moderating effect of the product type. Thus, we find no evidence that the impact of expressed uncertainty differs across product type.

Considering the reviewer-related signals, we observe that reviews provided by experienced reviewers are more helpful than reviews contributed by less-experienced reviewers. This relationship is significant at a 0.1% level, which supports H4. Furthermore, disclosing a reviewer's real name has a significant influence on review helpfulness, but the influence is negative instead of positive (H5 rejected).

Finally, related to the control variables, we observe that online reviews with star ratings that deviate from the average rating are evaluated as less

Table 5
Tobit-regression estimates explaining review helpfulness (n = 12,330).

Hypothesis		Variable	(1) Base model		(2) +Content		(3) +Reviewer	
			Coef.	p-Value	Coef.	p-Value	Coef.	p-Value
		Constant	1.3350	<0.001***	1.2438	<0.001***	2.2345	<0.001***
Review content-related Signals	H1	Quality			0.7820	<0.001***	0.4873	0.004**
		Quality × ProductType			−0.4301	0.094	−0.1947	0.413
	H2	DirectionalSentiment			0.6168	<0.001***	0.4568	<0.001***
		DirectionalSentiment × ProductType			−1.6269	<0.001***	−1.4262	<0.001***
	H3	Uncertainty			−0.6014	0.023*	−0.8349	0.001***
		Uncertainty × ProductType			−0.0764	0.809	0.2962	0.311
Reviewer-related signals	H4	Rank					−0.0450	<0.001***
	H5	RealName					−0.0180	<0.001***
Control variables		ProductType	0.0583	<0.001***	0.1830	<0.001***	0.1205	<0.001***
		Extremity	−0.1221	<0.001***	−0.1162	<0.001***	−0.0881	<0.001***
		Extremity × ProductType	−0.0712	<0.001***	−0.0727	<0.001***	−0.0712	<0.001***
		Depth	0.0004	<0.001***	0.0005	<0.001***	0.0003	<0.001***
		Depth × ProductType	−0.0003	<0.001***	−0.0004	<0.001***	−0.0002	<0.001***
		$Depth^2$	0.0000	<0.001***	0.0000	<0.001***	0.0000	<0.001***
		$Depth^2$ × Product Type	0.0000	<0.001***	0.0000	<0.001***	0.0000	<0.001***
		Age	0.0000	<0.001***	0.0000	<0.001***	0.0000	<0.001***
		TotalVotes	0.0000	0.010**	0.0000	0.012*	0.0000	0.016*
		Readability	−0.0001	0.651	−0.0004	0.131	−0.0004	0.071
		Product controls	Included		Included		Included	
		Monthly controls	Included		Included		Included	
		$p > \chi^2$		<0.001***		<0.001***		<0.001***
		Pseudo R^2		0.725		0.772		0.973
		Δ Pseudo R^2				+0.047		+0.201

*/**/*** indicate significance at a 5%/1%/0.1% level.

helpful. This influence is significant at a 0.1% level. Furthermore, the control variables for product type, review depth, squared review depth, review age, total number of votes and the moderating effects significantly affect review helpfulness. Focusing on the monthly dummy variables, we find that online reviews contributed in August are less helpful than online reviews written in other months. This might be explained by the fact that reviewers spend less effort when writing online reviews during their vacation.

We also observe that the hypothesis suggesting there is no influence of the different independent variables on review helpfulness can be rejected at the 0.1% level of significance. Furthermore, as shown by the difference regarding the Pseudo R^2 measure, we observe that the review content-related signals, and especially the reviewer-related signals, add explanatory value. Here, reviewer-related signals have an increased explanatory power on review helpfulness, which can be driven by the fact that they are considered more reliable and are thus more efficient regarding the reduction of uncertainty.

4.2.2. Influence of the signaling environment

In order to evaluate the influence of the signaling environment on signal processing, we analyze which effects prevail in case of (1) a low information environment as well as (2) a high information environment (Table 6). We therefore re-ran our analyses on a sample only containing those online reviews for products with up to ten online reviews (Model 1, "low information environment", since ten reviews can regularly be presented on one page). Furthermore, we re-run our analyses also on a sample containing only those online reviews on products with >10 reviews (Model 2, "high information environment"). When other thresholds are chosen (i.e. 5 or 25 reviews), the results remain robust.

We find that the observed effects prevail in a high information environment, as already outlined in the previous section. However, the influence of product quality as well as uncertainty vanishes in the case of a low information environment (H6 confirmed). Consequently, the signaling environment has an influence on signal processing, whereas in a high information environment, specific signals are required in order to make the online review helpful.

4.2.3. Influence of signaler incentives

Table 7 shows whether the signals sent via incentivized and non-incentivized online reviews differ. Therefore, we compare online reviews which have been written after the reviewer has received a free product with the sample of non-incentivized online reviews.

The results show a significant difference concerning product quality and strength of review sentiment. Consequently, those reviewers having received free products specifically take care that their online reviews

Table 7
Signal provision in case of incentivized and non-incentivized reviews.

Variable		Incentivized reviews	Non-incentivized reviews	
		Mean	Mean	p-value
Review content-related signals	Quality	0.0226	0.0181	<0.001***
	DirectionalSentiment	0.0552	0.0627	0.014*
	Uncertainty	0.0186	0.0174	0.518
Reviewer-related signals	Rank	7.9729	13.7941	<0.001***
	RealName	0.5082	0.3591	<0.001***
	n	368	11,975	

*/**/*** indicate significance at a 5%/1%/0.1% level.

contain signals which are helpful for purchase decision making in case of product quality (significant at a 0.1% level), but use less emotional language (significant at a 5% level). Furthermore, incentivized online reviews are contributed by reviewers with a higher rank and by reviewers disclosing their real name more frequently. Thus, signaler incentives have an impact on signaling (H7 confirmed).

4.3. Predictive performance

As shown by the explanatory Tobit analysis, the proposed research model explains a significant variance in review helpfulness. To investigate whether the model is useful for online retailers and can be applied to properly predict the helpfulness of previous unrated online product reviews, we estimate the model on the training sample and validate the model by means of a simulation based on the holdout-sample. This ensures realistic evaluation results that avoid the risk of overfitting [51,65]. Therefore, we compute classic evaluation metrics (i.e., predictive accuracy and the correlation between the actual and predicted rank) and present the results of our practical evaluation scenario in which an online retailer aims at predicting and displaying the most-helpful product reviews.

We compare the results of different models: (1) a ranking based on the model proposed within this paper, (2) a ranking based on random ordering, (3) a ranking based on the time of review (older reviews are ranked better), (4) a ranking based on the seminal models proposed by Forman et al. [2] and (5) Mudambi and Schuff [4], and (6) a ranking based on the recent model by Salehan and Kim [66]. We have selected these three models because they emphasize different signal categories and consequently different factors explaining the helpfulness of online product reviews. While Forman et al. [2] incorporate factors related to both the review (equivocality) and the reviewer (self-disclosure), Mudambi and Schuff [4] present a model with a special focus on review characteristics (review extremity and review word count). Finally,

Table 6
Research model in case of (1) low and (2) high information environment.

Hypothesis		Variable	(1) Low information environment		(2) High information environment	
			Coef.	p-value	Coef.	p-value
		Constant	2.3281	<0.001***	2.2570	<0.001***
Review content-related signals	H1	Quality	0.6456	0.196	0.4462	0.014*
		Quality × ProductType	−0.8062	0.292	−0.2023	0.419
	H2	DirectionalSentiment	1.2212	<0.001***	0.3398	0.001***
		DirectionalSentiment × ProductType	−0.9182	0.019*	−1.3520	<0.001***
	H3	Uncertainty	−0.7081	0.378	−0.8703	0.001***
		Uncertainty × ProductType	1.0306	0.318	0.2727	0.371
Reviewer-related signals	H4	Rank	−0.0383	<0.001***	−0.0452	<0.001***
	H5	RealName	−0.0093	0.548	−0.0192	<0.001***
Control variables		Control variables	Included		Included	
		p > χ^2		<0.001***		<0.001***
		Pseudo R^2		0.7294		1.0254

*/**/*** indicate significance at a 5%/1%/0.1% level.

Table 8
Model comparison based on classic evaluation metrics.

	Accuracy	Rank correlation
Proposed model	81.09%	0.8308
Random	–	0.6312
Time-based	–	0.5904
Forman et al. [2]	57.46%	0.6148
Mudambi and Schuff [4]	68.48%	0.8513
Salehan and Kim [66]	61.20%	0.6826

Salehan and Kim [66] emphasize more-specific review content characteristics and incorporate review length and sentiment in order to explain helpfulness.

Table 8 shows the results of the classic evaluation metrics. Focusing on predictive accuracy and showing the proportion of true results among the total number of cases examined, we observe that the proposed model clearly outperforms the other models, being able to correctly classify 81.09% of all cases. Focusing on the rank correlation (i.e., the correlation between the actual and the predicted rank), we observe that the proposed model has a good performance as well and outperforms most of the other benchmark models. Nevertheless, the model proposed by Mudambi and Schuff [4] has a slightly higher rank correlation.

In Table 9, we focus on our problem-specific evaluation scenario and find that the mean absolute error when determining the most-helpful online product review based on the model proposed within this study is 4.4658. Thus, the application of the proposed model leads to an error of approximately 4 ranks (i.e., 4 online product reviews are ranked better than the actual most-helpful one). Therefore, the proposed model clearly outperforms the benchmark models. For instance, if the model by Mudambi and Schuff [4] is used for ranking online product reviews, the mean absolute error is 8.5528. A Wilcoxon-signed-rank test for equality of the different models' errors shows that these are different at the 0.1% level of significance.

These results are also valid if more reviews than the most-helpful online product review are identified. For instance, the proposed model is also superior to the benchmark models when identifying the two or three most-helpful online product reviews. The results also hold true in comparison to the time-based and random rankings. In the case of the model recently proposed by Salehan and Kim [66], we also observe a higher mean absolute error when compared to the proposed model. Nevertheless, the difference is only significant for identifying the two and three most-helpful online reviews.

The different evaluations show that the proposed model is of high practical relevance, as it can be utilized to make superior predictions of online review helpfulness in the vast majority of cases. Furthermore, we observe that the proposed practical evaluation scenario is valuable, especially when compared to classical evaluation metrics. For instance, we can observe that although the model proposed by Mudambi and Schuff [4] has a slightly higher rank correlation than the proposed model, it

performs significantly worse with regards to identifying the most-helpful online reviews. We also observe that the model by Salehan and Kim [66] shows a clearly lower performance with respect to rank correlation and predictive accuracy. Nevertheless, when focusing on predicting the most-helpful online reviews, the results are more promising.

4.4. Discussion

Our results make clear that both suggested signal categories are relevant in the context of online product reviews. Regarding review-related signals, our results show that a deeper content analysis focusing on signals expressed in the review content provides further insights related to review helpfulness, specifically compared to basic text characteristics, such as a review's length. We can confirm the influence of signaled product quality, review sentiment, and review uncertainty on online review helpfulness.

We also observe that reviewer-related signals have an impact on online review helpfulness. Here, the reviewer-related information in the form of the user rank represents a signal regarding the reviewer's ability to write useful reviews, and thus positively influences review helpfulness.

Interestingly, the impact of disclosing the real name on review helpfulness is the opposite of what is expected. We also find that disclosing the real name has a negative impact on review helpfulness and does not signal higher reviewer credibility. One possible explanation for this result may be that reviewers disclosing their real name do not necessarily reveal their real opinion about the product. This explanation is supported by results from previous research where it is shown that feedback differs if it is given anonymously instead of non-anonymously [67]. It can also be argued that non-anonymity can be regarded as a signal that increases uncertainty and thus decreases helpfulness. Nevertheless, the question of whether and how anonymity affects reviewers should be further investigated in future research.

Comparing review content-related signals with reviewer-related signals, we observe that reviewer-related signals have a higher impact on the explanatory power than review content-related signals. This can be explained by the fact that the reviewer-related signals included in this study represent assessment signals which are seen as more reliable than the conventional content-related signals.

Our results also verify and broaden the insights about the product type's influence on the helpfulness of online reviews. Consistent with previous studies, we observe that reviews about search goods are more helpful than reviews related to experience goods [4]. Here, we also find that the different independent variables can have a different impact on the perceived helpfulness of online reviews across product categories. Specifically, we find that the strength of sentiment increases review helpfulness in the case of search goods while it decreases review helpfulness in the case of experience goods.

Considering the control variables, we verify previous research related to the negative influence of review extremity as well as the

Table 9
Mean Absolute Error (MAE) for predicting the helpfulness of the top n online product reviews.

Helpfulness rank	MAE most helpful		MAE two most helpful		MAE three most helpful	
	Mean	Median	Mean	Median	Mean	Median
Proposed model	4.4658	1	5.9421	1	5.8568	1
Random	10.1367	2	13.5306	2	14.1686	2
vs. Proposed Model		0.000***		0.000***		0.000***
Time-Based	14.9255	3	15.6720	2	16.1155	2
vs. Proposed Model		0.000***		0.000***		0.000***
Forman et al. [2]	11.1118	2	13.6142	2	12.8868	2
vs. Proposed Model		0.000***		0.000***		0.000***
Mudambi and Schuff [4]	8.5528	2	9.4727	2	11.1801	2
vs. Proposed Model		0.000***		0.000***		0.000***
Salehan and Kim [66]	5.3975	1	7.2797	2	9.2055	2
vs. Proposed Model		0.1367		0.000***		0.000***

*/**/*** indicate significance at a 5%/1%/0.1% level.

positive impact of review depth on review helpfulness. Mudambi and Schuff [4] observe a negative influence of the total number of votes on the helpfulness of online reviews. In contrast, we observe a small positive influence, because we consider online reviews that were voted on at least ten times to strengthen the reliability of our study.

Related to the influence of the signaling environment, we clearly observe that signals have a differing impact on review helpfulness, depending on the amount of further online reviews being publicly available. We confirm that in a low information environment, the existence of the online review itself is helpful, whereas in a high information environment, an online review has to specifically contain signals related to product quality and sentiment in order to distinguish itself from other reviews and to be helpful during the purchase decision making process. Focusing on the provisioning of reviewer incentives, we can confirm that reviewer incentives have an impact on the provisioning of signals. Here, reviewers specifically take care that their online reviews contain signals related to product quality as they have a positive influence on review helpfulness. Furthermore, these reviewers use less emotional language when composing their online reviews.

Related to the practical relevance indicated by our predictive evaluation, we demonstrate that our proposed model nearly halves the mean absolute error in comparison to the benchmark models when predicting review helpfulness. This shows that the results of this study are highly relevant for online retailers that aim at ranking online product reviews automatically (e.g., in case of new reviews that have not been voted on before, or in case of no possibility to assess helpfulness by means of user assessments).

Furthermore, we demonstrate the relevance of our novel evaluation scenario by evaluating and comparing the predictive performance of our proposed model with other benchmark models. Concentrating on predicting the most-helpful online product reviews reveals which models are specifically relevant for online retailers who want to display these reviews first. Here, we clearly show that our evaluation scenario provides more domain-specific insights compared to classic evaluation metrics.

Consistent with Mudambi and Schuff [4], the methodology applied to evaluate review helpfulness is subject to a limitation, as review helpfulness is quantified based on the votes of the readers who participated in the voting. Thus, this evaluation may not fully encompass the perceptions of customers who did not participate. Nevertheless, approximately 38 readers have evaluated a typical review analyzed within this study, so we are able to cover the perceptions of a substantial number of users. With regards to methodology, content analysis is subject to the limitation that the observations rely on the dictionary used. If a word characterizing a specific category is not contained in the dictionary, the results of the automated content analysis might be biased [54]. We mitigate this issue by making use of a standardized and well-established dictionary contained in the General Inquirer [58,59]. Additionally, in the field of sentiment analysis, approaches relying on term frequencies do not consider complex language concepts such as irony [68]. However, these constructs are oftentimes difficult to identify for individuals as well, and the term-based approaches (consistent with the approach used within our study) have already been utilized with success to analyze sentiment in other domains [69].

We use the Amazon rank to measure reviewer expertise. Due to the cross-sectional nature of this study, we are aware of the limitation that the rank is based on different factors, including a reviewer's previous review helpfulness score – also encompassing the actual review's helpfulness. Nevertheless, on average, reviewers within the sample have contributed >27 reviews, so the influence of a single review can be assumed to be small. Nevertheless, if an online retailer has recently introduced a helpfulness rating and is not yet able to determine a rank, our results show that textual content is also valuable for determining review helpfulness. Focusing on reviewer-related signals only bears the risk that reviewers may publish false information in their user profiles to appear in a positive light. However, because this study takes into account assessment signals, this risk can be assumed to be very low.

Finally, because online reviews have also been shown to influence sales, different market participants have already begun to make public very positive online reviews to boost the sales of their offered products or extremely negative reviews to reduce the turnover of their competitors [1]. As follows, an analysis of reviews posted on the Internet is accompanied by the risk that such fake reviews are included in the dataset, which might bias the results. Nevertheless, we analyze different products, so a manipulation of a single product or service would have only a minor impact on the results. Additionally, because the different products are best-sellers and are thus discussed within a large number of reviews, a potential manipulator would need to publish a large number of fake reviews, which makes manipulation time-consuming and, consequently, less probable.

5. Conclusion

Online product reviews have gained increased importance for online consumers as well as online retailers. A growing stream of literature investigates the factors explaining the helpfulness of online reviews. Our study builds upon signaling theory and presents two categories of signals that we incorporate in our research model on the drivers of review helpfulness. We therefore enhance the previous understanding as we provide signal categories relevant in the context of online reviews, whereas we find that review-related signals encompassing product quality, review sentiment, review uncertainty, and reviewer-related signals in the form of reviewer expertise and reviewer non-anonymity are relevant factors influencing review helpfulness. Here, the influence of reviewer-related signals is higher compared to review content-related signals. Furthermore, we contribute to signaling theory in two additional relevant aspects [20], as we observe that the signaling environment has an influence on signal processing and as we find that signaling incentives have an impact on signal provision. With a focus on prediction, our model enables online retailers to display the most-helpful reviews first. Our problem-specific evaluation scenario highlights the practical relevance of our model by demonstrating the predictive performance when predicting review helpfulness.

We show the high relevance of our results by means of a problem-specific evaluation scenario focusing on the prediction of review helpfulness. We also show that the proposed evaluation methodology provides more-specific assessments when compared to classic performance metrics.

This study also has important insights for online retailers. We provide insights on how to update the guidelines on how online reviews should be written so that readers perceive them to be helpful. Online retailers should generally advise their customers to describe product-related aspects, avoid uncertain language and express strong sentiment in the case of search goods and avoid sentiment in the case of experience goods. Online retailers might display such information on the web page where consumers can submit online product reviews. In addition, online retailers might also automatically predict the helpfulness of submitted reviews based on the proposed model. In case of low helpfulness, they might suggest changes to the reviewer before the review is finally published. Furthermore, the study shows that providing information about the review author next to the textual content is valuable as this information signals reviewer reliability.

Based on our results, there are multiple directions for further research. First, researchers can examine whether a reviewer's cultural background has an influence on the review style and the resulting helpfulness assessment or whether the product price is relevant for predicting review helpfulness. Furthermore, future research could use additional data sources such as tripadvisor.com or imdb.com to assess whether the results also hold for reviews discussing other spheres of interest, including hotels and movies. Finally, with the understanding gained from this study on how to select the most-helpful online reviews, researchers could also examine how many reviews are necessary for properly supporting purchase decisions and for avoiding information overload.

12

References

[1] J.A. Chevalier, D. Mayzlin, The effect of word of mouth on sales: online book reviews, J. Mark. Res. 43 (2006) 345–354.

[2] C. Forman, A. Ghose, B. Wiesenfeld, Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets, Inf. Syst. Res. 19 (2008) 291–313.

[3] F. Zhu, X. Zhang, Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics, J. Mark. 74 (2010) 133–148.

[4] S.M. Mudambi, D. Schuff, What makes a helpful online review? A study of customer reviews on amazon.com, MIS Q. 34 (2010) 185–200.

[5] A. Ghose, P.G. Ipeirotis, Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics, IEEE Trans. Knowl. Data Eng. 23 (2011) 1498–1512.

[6] D. Yin, S. Bond, H. Zhang, Dreading and ranting: the distinct effects of anxiety and anger in online seller reviews, ICIS 2011 Proceedings, 2011.

[7] L. Huang, C.-H. Tan, W. Ke, K.-K. Wei, Comprehension and assessment of product reviews: a review-product congruity proposition, J. Manag. Inf. Syst. 30 (2013) 311–343.

[8] P. Wu, H. van der Heijden, N. Korfiatis, The influences of negativity and review quality on the helpfulness of online reviews, ICIS 2011 Proceedings, 2011.

[9] N. Korfiatis, E. García-Bariocanal, S. Sánchez-Alonso, Evaluating content quality and helpfulness of online product reviews: the interplay of review helpfulness vs. review content, Electron. Commer. Res. Appl. 11 (2012) 205–217.

[10] J. Donath, Signals in social supernets, J. Comput.-Mediat. Commun. 13 (2008) 231–251.

[11] S.M. Kim, P. Pantel, T. Chklovski, M. Pennacchiotti, Automatically assessing review helpfulness, Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing 2006, pp. 423–430.

[12] M. Spence, Signaling in retrospect and the informational structure of markets, Am. Econ. Rev. 92 (2002) 434–459.

[13] M. Spence, Job market signaling, Q. J. Econ. 87 (1973) 355–374.

[14] M. Spence, Market Signaling: Informational Transfer in Hiring and Related Screening Processes, Harvard University Press, Cambridge, MA, USA, 1974.

[15] G.A. Akerlof, The market for "lemons": quality uncertainty and the market mechanism, Q. J. Econ. 84 (1970) 488–500.

[16] P.A. Pavlou, H. Liang, Y. Xue, Understanding and mitigating uncertainty in online exchange relationships: a principal-agent perspective, MIS Q. 31 (2007) 105–136.

[17] A. Celani, P. Singh, Signaling theory and applicant attraction outcomes, Pers. Rev. 40 (2011) 222–238.

[18] C. Lampe, N. Ellison, C. Steinfield, A familiar Face(book): profile elements as signals in an online social network, Proceedings of the Computer/Human Interaction Conference San Jose, California, USA, 2007.

[19] M. Siering, J. Muntermann, How to identify tomorrow's most active social commerce contributors? Inviting starlets to the reviewer hall of fame, Proceedings of the 34th International Conference on Information Systems, Milan, Italy, 2013.

[20] B.L. Connelly, S.T. Certo, R.D. Ireland, C.R. Reutzel, Signaling theory: a review and assessment, J. Manag. 37 (2011) 39–67.

[21] D. Grewal, J. Gotlieb, H. Marmorstein, The moderating effects of message framing and source credibility on the price-perceived risk relationship, J. Consum. Res. 21 (1994) 145–153.

[22] T.B. White, Consumer trust and advice acceptance: the moderating roles of benevolence, expertise, and negative emotions, J. Consum. Psychol. 15 (2005) 141–148.

[23] A.H. Eagly, S. Chaiken, An attribution analysis of the effect of communicator characteristics on opinion change: the case of communicator attractiveness, J. Pers. Soc. Psychol. 32 (1975) 136–144.

[24] V.A. Zeithaml, Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence, J. Mark. 52 (1988) 2–22.

[25] J.C. Sweeney, G.N. Soutar, Consumer perceived value: the development of a multiple item scale, J. Retail. 77 (2001) 203–220.

[26] J.A. Howard, J.N. Sheth, The Theory of Buyer Behavior, Wiley, New York, 1969.

[27] P. Nelson, Information and consumer behavior, J. Polit. Econ. 78 (1970) 311–329.

[28] C. Park, T.M. Lee, Information direction, website reputation and eWOM effect: a moderating role of product type, J. Bus. Res. 62 (2009) 61–67.

[29] I. Wen, Factors affecting the online travel buying decision: a review, Int. J. Contemp. Hosp. Manag. 21 (2009) 752–765.

[30] Z. Yang, X. Fang, Online service quality dimensions and their relationships with satisfaction: a content analysis of customer reviews of securities brokerage services, Int. J. Serv. Ind. Manag. 15 (2004) 302–326.

[31] J.J. McCluskey, A game theoretic approach to organic foods: an analysis of asymmetric information and policy, Agric. Resour. Econ. Rev. 29 (2000) 1–9.

[32] M.W. Uhl, Explaining U.S. consumer behavior with news sentiment, ACM Transactions on Management Information Systems, 2, 2011, pp. 1–18.

[33] Merriam-Webster, Merriam-Webster Online Dictionary: Sentiment, http://www.merriam-webster.com/dictionary/sentiment 2012.

[34] R.M. Schindler, B. Bickart, Perceived helpfulness of online consumer reviews: the role of message content and style, J. Consum. Behav. 11 (2012) 234–243.

[35] J.A. Krosnick, D.S. Boninger, Y.C. Chuang, M.K. Berent, C.G. Camot, Attitude strength: one construct or many related constructs? J. Pers. Soc. Psychol. 65 (1993) 1132–1151.

[36] G. Marks, N. Miller, The effect of certainty on consensus judgments, Personal. Soc. Psychol. Bull. 11 (1985) 165–177.

[37] P.D. Bennett, G.D. Harrell, The role of confidence in understanding and predicting buyers' attitudes and purchase intentions, J. Consum. Res. 2 (1975) 110–117.

[38] T. Mavlanova, R. Benbunan-Fich, M. Koufaris, Signaling theory and information asymmetry in online commerce, Inf. Manag. 49 (2012) 240–247.

[39] K.K. Kuan, K.-L. Hui, P. Prasarnphanich, H.-Y. Lai, What makes a review voted? An empirical investigation of review voting in online review systems, J. Assoc. Inf. Syst. 16 (2015) 48–71.

[40] P.B. Goes, M. Lin, C.-m.A. Yeung, "Popularity effect" in user-generated content: evidence from online product reviews, Inf. Syst. Res. 25 (2014) 222–238.

[41] T. Hennig-Thurau, K.P. Gwinner, G. Walsh, D.D. Gremler, Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the Internet? J. Interact. Mark. 18 (2004) 38–52.

[42] D.J. Kim, D.L. Ferrin, H.R. Rao, A trust-based consumer decision-making model in electronic commerce: the role of trust, perceived risk, and their antecedents, Decis. Support. Syst. 44 (2008) 544–564.

[43] M.L. Jensen, J.M. Averbeck, Z. Zhang, K.B. Wright, Credibility of anonymous online product reviews: a language expectancy perspective, J. Manag. Inf. Syst. 30 (2013) 293–324.

[44] T. Ong, M. Mannino, D. Gregg, Linguistic characteristics of shill reviews, Electron. Commer. Res. Appl. 13 (2014) 69–78.

[45] Q. Gan, Q. Cao, D. Jones, Helpfulness of online user reviews: more is less, AMCIS 2012 Proceedings, 2012.

[46] B. Jiang, J.A. Belohlav, S.T. Young, Outsourcing impact on manufacturing firms' value: evidence from Japan, J. Oper. Manag. 25 (2007) 885–900.

[47] S.A. Zahra, I. Filatotchev, Governance of the entrepreneurial threshold firm: a knowledge-based perspective, J. Manag. Stud. 41 (2004) 885–897.

[48] J.J. Janney, T.B. Folta, Moderating effects of investor experience on the signaling value of private equity placements, J. Bus. Ventur. 21 (2006) 27–44.

[49] J. Pu, Y. Kwark, S. Han, B. Gu, Q. Ye, The double-edged sword of expert reviewer programs: the effects of offering expert reviewer status on review generation, ICIS 2017 Proceedings, 2017.

[50] W. Shen, J. Rees Ulmer, Competing for attention: an empirical study of online reviewers' strategic behavior, MIS Q. 39 (2015).

[51] J. Han, M. Kamber, Data Mining: Concepts and Techniques, 2nd ed. Elsevier; Morgan Kaufmann, San Francisco, 2006.

[52] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, Quebec, Canada, 14, 1995.

[53] R.P. Weber, Measurement models for content analysis, Qual. Quant. 17 (1983) 127–149.

[54] S.D. Rosenberg, P.P. Schnurr, T.E. Oxman, Content analysis: a comparison of manual and computerized systems, J. Pers. Assess. 54 (1990) 298–310.

[55] P.P. Schnurr, S.D. Rosenberg, T.E. Oxman, G. Tucker, A methodological note on content analysis: estimates of reliability, J. Pers. Assess. 50 (1986) 601–609.

[56] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press, New York, NY, 2015.

[57] K. Krippendorff, Content Analysis: An Introduction to Its Methodology, Sage, Los Angeles, 2013.

[58] P.J. Stone, E.B. Hunt, A computer approach to content analysis: studies using the general inquirer system, Proceedings of the AFIPS Spring Joint Computer Conference 1963, pp. 241–256.

[59] P.J. Stone, R.F. Bales, J.Z. Namenwirth, D.M. Ogilvie, The general inquirer: a computer system for content analysis and retrieval based on the sentence as a unit of information, Behav. Sci. 7 (1962) 484–498.

[60] R.P. Weber, Basic Content Analysis, 2nd ed. Sage, Newbury Park, California, USA, 1990.

[61] E.F. Kelly, P.J. Stone, Computer Recognition of English Word Senses, North-Holland Publishing, Amsterdam/Oxford, 1975.

[62] Amazon, Amazon's top customer reviewers — how ranking works, http://www.amazon.com/gp/customer-reviews/guidelines/top-reviewers.html , Accessed date: 2 May 2017.

[63] E. Wolff-Mann, When did a 4-star review become a bad review?, Time, http://time.com/money/page/online-reviews-trust-fix/ , Accessed date: 15 August 2017.

[64] F.X. Diebold, Elements of Forecasting, South-Western College Pub, Mason, Ohio, 1998.

[65] D.M. Hawkins, The problem of overfitting, J. Chem. Inf. Comput. Sci. 44 (2004) 1–12.

[66] M. Salehan, D.J. Kim, Predicting the performance of online consumer reviews: a sentiment mining approach to big data analytics, Decis. Support. Syst. 81 (2016) 30–40.

[67] R. Lu, L. Bol, A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback, J. Interact. Online Learn. 6 (2007) 100–115.

[68] S. Baccianella, A. Esuli, F. Sebastiani, Multi-facet rating of product reviews proceedings of the ECIR 2009, LNCS 5478 (2009) 461–472.

[69] P.C. Tetlock, Giving content to investor sentiment: the role of media in the stock market, J. Financ. 62 (2007) 1139–1168.

Michael Siering is a postdoctoral research associate at Goethe University Frankfurt. His research focuses on decision support systems in electronic markets, with a focus on the analysis of user generated content. His work has been published in journals such as Decision Support Systems, Journal of Management Information Systems, and Journal of Information Technology.

Jan Muntermann is a professor and chair of electronic finance and digital markets in the faculty of business and economics at the University of Goettingen. His research interests include digital business strategy development and execution, business intelligence and analytics, and research methodology. He has published in journals such as Information Systems Research, Decision Support Systems, Information & Management and the European Journal of Information Systems.

Balaji Rajagopalan is the Dean of the College of Business at Northern Illinois University. Rajagopalan's expertise is in how firms and communities leverage information technologies. His research has been published in journals such as Information Systems Research, Information Systems Journal, and European Journal of Operational Research.